ChinaXiv合作期刊

干旱区地理
ARID LAND GEOGRAPHY

陕西省人口分布影响因素的空间计量分析①

米瑞华[1,2]， 高向东[1]
（1 华东师范大学公共管理学院，上海 260002； 2 延安大学经济与管理学院，陕西 延安 716000）

摘 要：人口分布影响因素研究有利于揭示人口分布规律，预判人口分布趋势。基于陕西省区县级人口、经济社会、自然地理等数据，通过因子分析方法和空间计量建模解析人口分布的影响因素。研究发现，人口地域别比率不仅取决于一个特定区县内可观测的经济社会、历史基础、自然地理等外在特征，还取决于该区县不可观测的、模型遗漏的其他共有特征，其中经济与公共服务因子、人口基底因子对人口分布具有最显著的正向解释力，其他因素影响较弱或统计不显著；城市等级可显著强化产业结构、人均收入和地形因素对人口分布的影响。研究认为，经济与公共服务因素是优化人口分布的关键，同时需考虑自然地理因素的限制作用。研究对人口分布优化政策的制定具有参考价值。

关 键 词： 人口分布；影响因素；人口地域别比率；空间计量模型；陕西省
文章编号：

chinaXiv:202006.00008v1

人口分布影响因素研究对认识人口分布规律、预判人口分布变化趋势具有重要参考价值，可作为人口分布预测、调控的科学依据[1]。区域人口分布与其自然地理、历史积淀、经济机遇、社会发展、行政等级等因素有密切联系，且影响因素之间的交互作用非常复杂。三因素理论等将影响人口分布的因素分为三大类[2]：自然环境因素、经济发展因素和社会发展因素。经典的“推—拉”理论认为充足的就业机会，较高的工资收入，良好的教育条件、生活条件等会促进区域人口迁入和净增长。新古典经济学的宏观和微观迁移理论基于即期或预期的净收益来解释人口流动及人口分布变动[3]。

大量实证研究关注人口分布的影响因素，基于省级[4-6]、地市级[7]、区县级[8-10]、乡镇级[11]等不同的空间尺度，采用主成分分析法[12]、空间叠加耦合分析[8]、相关分析[4,13]、探索性空间分析[5,14]、空间关联分析[9]、多元线性回归方法[11,15-17]、偏最小二乘法[18]、地理加权回归[10,13,19]、面板模型[7]等多种方法，对人口分布的影响因素进行实证研究，具体指标的选取包括自然、经济、社会、政治、科教、文化、历史等各个方面[1,4]。孟向京等[4]发现自然因素对中国人口分布影响最大，经济因素次之。张耀军等[19]认为社会经济因素对毕节地区人口分布的影响大于自然地理因素，其中综合经济实力、城镇化水平、交通条件和医疗条件对人口密度正相关，地形条件与人口密度负相关。王露等[18]发现人口规模在人口下降地区存在推力，高经济发展水平在人口密度增加地区存在拉力。沈思连等[10]认为人均 GDP 和人均粮食产量是影响河南省人口分布的主要因素。汪思言等[13]研究表明珠江流域人口密度与多年平均降雨呈现正相关，地形条件和交通条件是控制因素。王利等[6]发现综合经济实力、农业因素和地形条件是影响河北省人口分布的重要原因。已有成果取得较大进展，但仍存在改进空间。首先，部分研究成果对模型是否满足假定前提的检验不足，影响了分析结果的信度。其次，很多学者采用逐步回归法，统计上不显著的变量被剔除，有可能导致遗漏关键变量和模型误设。最后，根据地理学第一定律，地理事物或属性在空间分布上往往存在空间依赖性或空间异质性[20]，当存在空间自相关时应采用空间计量分析方法，但目前应用空间计量模型实证人口分布影响因素的研究相对较少[7]。

① **收稿日期：**2018-10-31； **修订日期：**2019-01-12
基金项目：国家社科基金重点项目(18ARK001)；国家社科基金重点项目(19ARK004)；陕西省科技厅项目(18JK0850)；延安市社科专项规划项目(19BDD08)资助
作者简介：米瑞华(1983-)，博士后，副教授. E-mail：rice221@163.com
通讯作者：高向东(1963-)，博导，教授. E-mail：xdgao@ecnu.edu.cn

ChinaXiv合作期刊

chinaXiv:202006.00008v1

陕西省位于中国西北内陆腹地,包括陕北高原生态脆弱区、关中平原政治文化经济中心区、以及陕南秦巴山地国家生态屏障区[21],辖西安、宝鸡等 10 个地级以上城市,107 个区县。2015 年末总人口为 3.79×10^7 人,省内人口密度最高的碑林区 2.75×10^4 人·km^{-2},最低的黄龙县 18.07 人·km^{-2},人口空间分异显著[22]。陕西省南北差异大,土地利用类型多样[23],县域经济发展差距较大,公共服务和基础设施分布不均等。揭示陕西人口空间分异的影响因素具有研究的典型性。

1 指标选取和数据来源

1.1 人口分布指标

衡量人口分布的指标主要有人口密度、人口地域别比率、人口集中指数(又称胡佛指数)等[1,24]。已有研究大多采用人口密度来刻画人口分布格局[7-8,11,19],也有少数学者使用人口规模[10]、人口地理集中度[17]作为衡量指标。

陕西省下辖区县的行政面积差异较大(面积最大的神木县 7.48×10^3 km^2,最小的碑林区仅 22.9 km^2)[23],以人口密度衡量人口分布容易导致误差项随空间单元面积变化而变化的严重异方差问题,影响回归分析的信度。人口地域别比率是反映人口分布格局的指标之一,是各空间单元对全部地域总人口的比例[24],计算公式为:

$$RPP_i = \frac{POP_i}{\sum POP_i} \times 100\% \tag{1}$$

式中:RPP_i 为 i 区县人口地域别比率;POP_i 为 i 区县人口数;$\sum POP_i$ 为全部地域总人口数。RPP_i 有效排除了行政面积差异的影响,反映了各区县人口在全部地域人口中所占份额和比例关系。

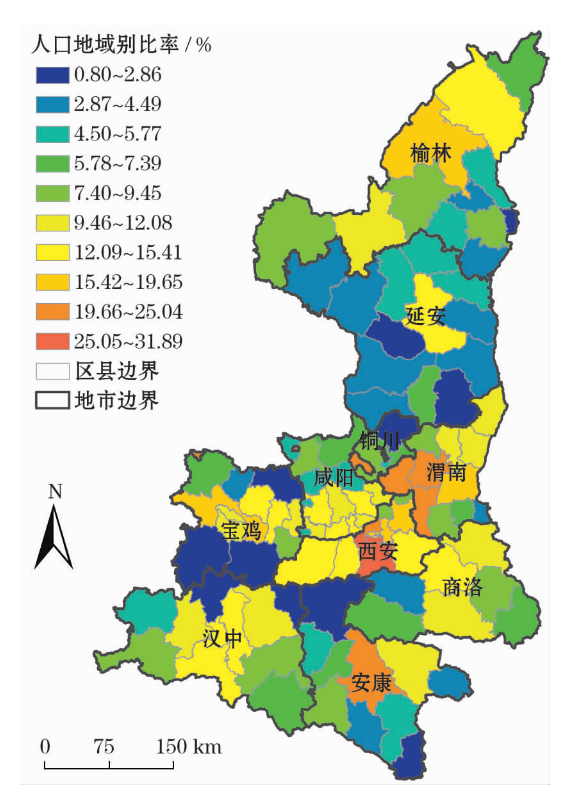
经计算,陕西省人口地域别比率最高的雁塔区人口占全省总人口的 31.89‰,最低的佛坪县只占全省人口的 0.08‰,省会城市核心区、地级以上城市以及关中平原各区县人口地域别比率较高,而地形崎岖、海拔较高且经济发展相对落后区县则较低(图 1)。

1.2 解释变量相关指标及因子分析降维

解释变量的选择基于三因素理论、推拉理论等,考虑时效性和可获取性等原则[2],初步选取 25 个经济社会发展指标,6 个自然地理指标。其中,2015 年陕西省人口、经济数据来源于《陕西区域统计年鉴 2016》;陕西省县级行政区划矢量地图数据来源于国家科技基础条件平台 – 国家地球系统科学数据共享服务平台(http://www.geodata.cn);地形起伏度、海拔等自然地理指标来源于中国寒区旱区科学数据中心(http://westdc.westgis.ac.cn),由"中国 1 km 分辨率数字高程模型数据集"提取、计算、分析获得。原始变量的描述性统计性质见表 1。

原始变量较多且存在信息冗余。为了避免主观选择变量造成遗漏变量或严重多重共线性,对其进行因子分析。原始变量的 KMO 值为 0.871,接近于 1;Bartlett 球形检验的近似卡方值为 3 770.841,说明适合采用因子分析降维。将原始指标分为经济社会发展类和地理类两组,采用主成分分析法,基于特征值大于 1 抽取主成分。

经济社会发展类变量最终保留 4 个有效主成分,分别反映了经济社会类总体信息的 51.800%、17.627%、12.205% 和 5.099%,累计反映了总体特征的 86.732%。根据因子载荷命名为"经济与公共服务因子"、"人口基底因子"、"产业结构因子"和"人均收入因子"。其中"经济与公共服务"得分越高,表示经济总量越大、公共服务越好。"人口基



注:审图号为陕 S(2018)007 号

图 1 2015 年陕西省人口地域别比率

Fig. 1 Regional proportion of population in Shaanxi Province in 2015

ChinaXiv合作期刊

chinaXiv:202006.00008v1

表 1 原始变量的描述性统计

Tab. 1 Descriptive statistics of original variables

变量类型	原始变量	单位	均值	标准差
经济	生产总值	10^8 元	163.1	182.4
总量	全社会固定资产投资	10^8 元	173.9	172.2
	地方财政收入	10^8 元	8.1	11.6
农业	第一产业产值	10^8 元	14.9	10.1
	粮食产量	10^4 t	12.1	8.9
	农林牧渔业总产值	10^8 元	26.0	17.4
产业	第二产业产值	10^8 元	83.1	91.5
结构	第三产业产值	10^8 元	65.1	109.7
	二产产值占 GDP 比重	%	48.9	16.4
	三产产值占 GDP 比重	%	37.3	13.3
	工业产值占 GDP 比重	%	36.6	21.5
人均	人均生产总值	元	42 222.0	26 275.9
收入	城镇居民人均可支配收入	元	26 663.6	3 456.2
	农村居民人均纯收入	元	9 543.5	2 354.9
消费	社会消费品零售总额	10^8 元	61.5	116.3
	地方财政支出	10^8 元	22.4	10.8
教育	普通小学专任教师数	人	1 344.1	762.0
	普通小学在校学生数	人	21 336.3	15 625.1
	普通中学专任教师数	人	1 571.6	1 042.8
	普通中学在校学生数	人	17 245.6	12 052.7
医疗	卫生机构床位数	张	1 946.9	1 853.9
	卫生技术人员	人	2 492.3	2 612.9
	执业(助理)医师	人	786.7	906.4
	注册护师、护士	人	988.0	1 196.6
人口	2000 年底总人口	10^4 人	33.6	19.8
地理	高程最小值	m	565.1	236.4
	高程最大值	m	1 740.8	669.0
	高程均值	m	1 004.3	336.7
	地形起伏度	m	1 175.7	694.5
	高程离差	m	245.4	168.6
	高程方差	m	88 378.9	117 438.9

底"得分越高,表示农业基础越好,人口基数越大;"产业结构"得分越高,表示二产比重越高,三产比重越低;"人均收入"得分越高,表示城乡人均收入水平越高。

地理类原始变量最终保留 2 个有效主成分,分别反映了总体信息的 67.176% 和 28.326%,累计反映了地理信息总体特征的 95.502%。根据因子载荷命名为"地形起伏度因子"和"平均海拔因子"。其中"地形起伏度"得分越高,表示空间单元地形起伏度越大;"平均海拔"得分越高,表示空间单元平均海拔越高。

因子得分图显示了陕西省经济社会发展和自然地理分布的复杂性(图 2)。经济与公共服务发展较好的区县集中在省会城市和地级以上城市核心区以及部分工矿城市;农业基础优越、人口基底较大的区县主要分布在关中平原以及陕南汉中、安康等地;第二产业为主的区县主要分布在陕北能源化工基地、以及铜川、宝鸡等工业城市;人均收入较高的区县主要分布在陕北能源区和西安市区及近郊;地形起伏度较大的区县集中分布在陕南秦巴山区;平均海拔较高的区县主要分布在陕北榆林、延安以西以及秦岭脊线。可见,陕西省绝大多数区县具有各类因素非协调发展的特点。

2 OLS 建模与检验

根据三因素理论和"推—拉"理论,以人口地域别比率为被解释变量,以经济与公共服务、人口基底、产业结构、人均收入、地形起伏度、平均海拔等因子为解释变量,建立 OLS 回归模型如下:

$$RPP_i = \beta_0 + \beta_1 Eco_i + \beta_2 Bas_i + \beta_3 Indus_i + \beta_4 Incom_i + \beta_5 Topo_i + \beta_6 Alt_i \quad (2)$$

式中:下标 i 表示空间单元的样本编号。带入样本数据,OLS 模型拟合如下:

$$\underset{}{RPP_i} = \underset{(0.000)}{9.35} + \underset{(0.000)}{3.88} Eco_i + \underset{(0.000)}{4.05} Bas_i - \underset{(0.000)}{0.65} Indus_i + \underset{(0.003)}{0.68} Incom_i + \underset{(0.013)}{0.43} Topo_i - \underset{(0.013)}{0.39} Alt_i \quad (3)$$

经检验,JB 统计量拒绝了 OLS 残差服从正态分布的原假设,提示可能存在模型误设或遗漏了重要变量。残差图提示少数奇异样本点干扰了残差的正态性,且这些奇异点都是人口快速增长的省会、地级市核心区等经济热点地区。结合近年来人口城镇化及相关经济政策因素,考虑"城市等级"虚拟变量对人口分布有重要影响[25]。

加入"城市等级"虚拟变量的 OLS 模型拟合如下:

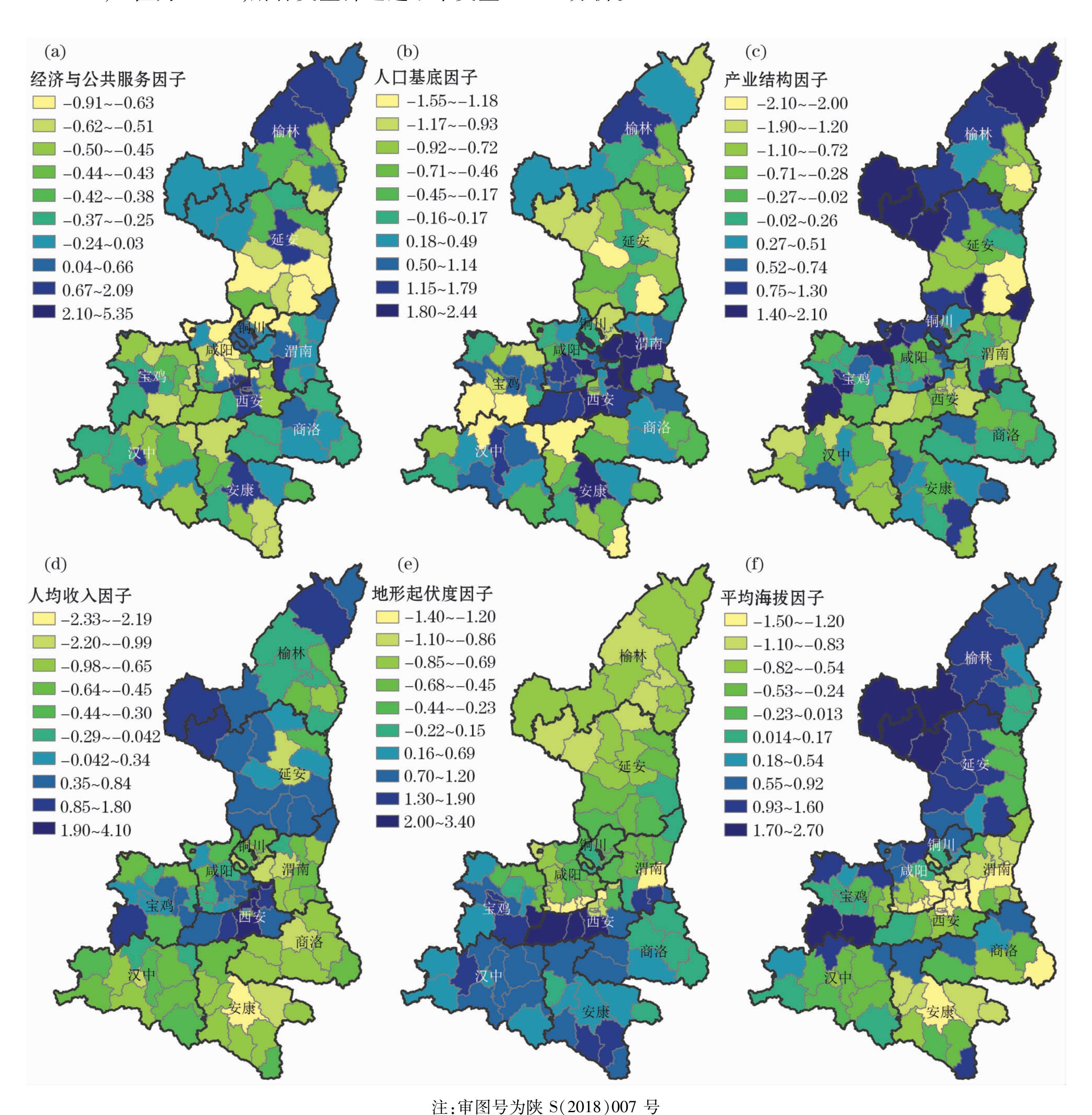
$$RPP_i = \underset{(0.000)}{9.42} + \underset{(0.000)}{4.13} Eco_i + \underset{(0.000)}{2.02} Eco_i \times D2_i + \underset{(0.000)}{3.96} Bas_i - \underset{(0.000)}{0.65} Indus_i - \underset{(0.001)}{1.12} Indus_i \times D1_i + \underset{(0.000)}{6.92} Indus_i \times D2_i + \underset{(0.000)}{1.28} Incom_i \times D1_i + \underset{(0.000)}{1.58} Topo_i \times D1_i - \underset{(0.000)}{0.42} Alt_i \quad (4)$$

式中:$D1 = 1$ 表示"地级以上城市核心区",$D1 = 0$ 表

ChinaXiv合作期刊

示“非地级市核心区”;$D2=1$ 表示“省会城市核心区”,$D2=0$ 表示“非省会城市核心区”。模型中不带虚拟变量的斜率系数反映了参照类(即 $D1=D2=0$)的斜率系数平均值,带有虚拟变量的斜率系数则体现了对应类与参照类在均值上的差。

加入虚拟变量后,JB 统计量为 2.37,P 值为 0.306,接受了残差服从正态分布的原假设;调整拟合优度 $\overline{R}^2$ 为 0.977;总体显著性检验 F 统计量为 509.543,P 值为 0.000;解释变量都通过了单变量显著性 t 检验;VIF 均小于 7.5,通过了多重共线性检验;BP 统计量为 11.929,P 值为 0.320,接受了同方差的原假设;DW 统计量为 1.736,大于 1% 显著性水平下的上限 $dU=1.65$,不存在序列自相关;被解释变量的 Moran's I 为 0.819,非常接近 1,OLS 残差的 Moran's I 为 0.139,P 值为 0.007,说明被解释变量和误差项都存在空间自相关,违背了高斯经典假定,OLS 回归将产生有偏估计,需采用空间计量模型分析。



注:审图号为陕 S(2018)007 号

图 2　陕西省经济社会、自然地理因子空间分布图

Fig. 2　Spatial distribution maps of socio-economic and physiologic factors of Shaanxi Province

chinaXiv:202006.00008v1

ChinaXiv合作期刊

3 空间计量建模与检验

空间计量建模技术可以有效处理变量或误差项存在空间自相关时 OLS 方法拟合系数的偏误。最常见的是空间滞后 SLM 模型和空间误差 SEM 模型[20]。模型形式:

空间滞后 SLM 模型强调被解释变量是空间自相关的,模型表达式为:

$$Y = \rho WY + X\beta + \varepsilon \quad (5)$$

$$\varepsilon = N(0, \delta^2 \ln)$$

式中:Y 为被解释变量;X 为 $n \times k$ 维的外生解释变量矩阵(n 为空间单元个数,k 为含常数项的解释变量个数);W 为 $n \times n$ 维空间权重矩阵;WY 为空间权重为 W 的被解释变量的空间滞后项;ρ 为空间滞后系数;β 为解释变量矩阵的回归系数;ε 为服从 0 均值、同方差、正态分布的随机误差向量。

空间误差 SEM 模型强调误差项是空间自相关的,模型表达式为:

$$Y = X\beta + \mu \quad (6)$$

$$\mu = \lambda W\mu + \varepsilon$$

$$\varepsilon \sim N(0, \delta^2 \ln)$$

式中:误差项 μ 由误差的空间滞后项 $W\mu$ 和独立正态分布的随机扰动项 ε 组成;λ 为空间误差自相关系数;其他参数与 SLM 模型参数含义相同。

在 GeoDa 软件中采用 Rook Contiguity 方法生成空间权重矩阵 W,并拟合 SLM 和 SEM 模型(表 2)。OLS 与 SLM、SEM 模型的截距和斜率系数符号一致,拟合优度均较高,解释变量都通过 t 检验,SLM 模型的空间滞后系数 和 SEM 模型的空间误差系数 λ 均在 10% 的显著性水平上拒绝了无异于零的原假设。

为了确定最佳模型,基于 OLS 残差进行 LM 诊断[20]。经计算,LM-*Lag* 值为 3.812,P 值为 0.051,LM-*Error* 值为 3.120,P 值为 0.077,都在 10% 的显著性水平上拒绝了不存在空间自相关的原假设[26-27]。Robust LM-*Lag* 统计量为 2.189,P 值为 0.139,认为因变量不存在空间自相关;Robust LM-*Error* 统计量为 6.932,P 值为 0.031,表明在 5% 的显著性水平下可以拒绝残差不存在空间自相关的原假设,说明 SEM 模型更适合[28]。AIC 准则也表明 SEM 模型优于其他模型[29]。

表 2 OLS 与空间计量回归结果

Tab. 2 Regression results of OLS and spatial econometric models

自变量	OLS 回归	空间滞后模型 SLM	空间误差模型 SEM
截距	9.42 (0.000)***	8.83 (0.000)***	9.40 (0.000)***
Eco_i	4.13 (0.000)***	4.10 (0.000)***	4.09 (0.000)***
$Eco_i \times D2_i$	2.02 (0.000)***	1.99 (0.000)***	2.05 (0.000)***
Bas_i	3.96 (0.000)***	3.95 (0.000)***	3.96 (0.000)***
$Indus_i$	-0.65 (0.000)***	-0.65 (0.000)***	-0.66 (0.000)***
$Indus_i \times D1_i$	-1.12 (0.001)***	-1.19 (0.000)***	-1.07 (0.001)***
$Indus_i \times D2_i$	6.92 (0.000)***	7.26 (0.000)***	6.95 (0.000)***
$Incom_i \times D1_i$	1.28 (0.000)***	1.14 (0.000)***	1.27 (0.000)***
$Topo_i \times D1_i$	1.58 (0.000)***	1.59 (0.000)***	1.52 (0.000)***
Alt_i	-0.42 (0.000)***	-0.33 (0.003)***	-0.42 (0.000)***
空间滞后(ρ)		0.06 (0.065)*	
空间误差(λ)			0.25 (0.061)*
拟合优度 R^2	0.979 2	0.979 9	0.980 3
似然函数对数	-136.52	-134.74	-134.63
AIC	293.05	291.48	289.27

注:***、**、*分别表示统计量在 1%、5% 和 10% 水平上显著

4 结果分析

空间误差 SEM 模型认为陕西省一个特定区县的人口地域别比率不仅取决于该区县一系列可观测的经济社会、历史基础、自然地理等外在特征,还取决于不可观测的、模型中遗漏的其他共有特征。其中,以“经济与公共服务因子”表征的经济社会发展情况对人口地域别比率有显著的正向解释力;以“人口基底因子”表征的农业基础条件和人口历史规模对人口地域别比率也有显著的正向解释力;以“产业结构因子”表征的二产比重较大(相应的三产比重较小)的产业结构特征对人口地域别比率有负向解释力,说明相对于服务业来说,第二产业对就业的吸纳能力有限;“平均海拔因子”对人口地域别比率有负向解释力。“人均收入因子”的影响统计上

chinaXiv:202006.00008v1

ChinaXiv合作期刊

chinaXiv:202006.00008v1

不显著的可能原因是陕西省人均收入较高的区县大多分布于能源资源矿产区;"地形起伏度因子"的影响统计上不显著的可能原因是人口分布同时受到降水、温度、湿度等自然环境的影响,而遗漏的环境变量与地形起伏度有较大的空间相关性,从而使地形起伏度因子的回归系数受到遗漏环境变量的反向作用而失去显著性。

"城市等级"显著改变各类影响因素对人口地域别比率作用的强度。同等经济与公共服务条件下,省会城市核心区比其他区县人口地域别比率平均高2.05‰;同等人均收入或地形起伏度条件下,地级市比一般区县的人口地域别比率更高;同等产业结构条件下,省会城市核心区较其他区县人口地域别比率更高,但地级市比一般区县人口地域别比率略低,说明省会城市比地级市有更强的产业结构调整适应性,而地级市若产业结构过于单一可能导致就业有限,因此更应优化产业结构。

5 结 论

(1)空间误差SEM模型是刻画陕西省人口分布影响因素的最优模型。模型揭示陕西省一个特定区县的人口地域别比率不仅取决于该区县一系列可观测的解释变量外在特征,还取决于其他不可观测的、从模型中遗漏的共有的特征。

(2)优化人口分布的关键可控指标是经济社会类变量,同时不可忽视自然地理的限制作用。当自然地理条件不变时,增大经济社会要素的空间布局集聚程度可吸引人口聚集,而降低经济社会要素的集聚程度可促进人口疏解。

(3)城市等级将显著改变各类影响因素对人口地域别比率作用的强度。较高行政等级的城市核心区由于历史、政策、自然条件等因素累积作用,集聚各类优质经济社会要素,导致城市等级越高,人口的聚集力越大。

人口分布影响因素在不同的研究区域,不同的数据来源、空间尺度、指标选择和分析方法,实证结论有所不同。本文通过因子分析避免了严重多重共线性对模型的不利影响,并通过考虑被解释变量和误差项的空间交互作用,纠正了OLS模型的有偏估计,以准确解析经济社会、自然地理等因素对人口分布的影响。其次,当行政面积差异较大且土地类型多样时,人口地域别比率指标较之于人口密度指标可有效缓解异方差等不利影响。

研究的局限性包括指标遴选、以及空间计量经济模型设定等方面。首先,指标中没有纳入分区县的降水、植被等自然环境指标;其次,因子分析方法的固有缺陷导致解释变量无量纲,只能通过相对值刻画其影响程度;最后,空间计量模型除了本文使用的经典的空间滞后和空间误差模型外,还有空间杜宾、空间杜宾误差等多种模型形式。未来可通过补充自然环境指标、改进指标遴选方法、设定含有更多空间效应的模型、使用空间面板数据等途径改进研究。

参考文献(References)

[1] 张耀军. 京津冀城市圈人口有序流动及合理分布[J]. 人口与发展,2015,(2):33-38. [ZHANG Yaojun. Orderly flow and reasonable distribution of population in Beijing,Tianjin and Hebei Urban Circle[J]. Population and Development,2015,(2):33-38.]

[2] 张善余. 中国人口地理[M]. 北京:科学出版社,2007:243-285. [ZHANG Shanyu. Population geography of China[M]. Beijing:Science Press,2007:243-285.]

[3] 李竞能. 现代西方人口理论[M]. 上海:复旦大学出版社,2004:139-179. [LI Jingneng. Modern western population theory[M]. Shanghai:Fudan University Press,2004:139-179.]

[4] 孟向京,贾绍凤. 中国省级人口分布影响因素的定量分析[J]. 地理研究,1993,12(3):56-63. [MENG Xiangjing,JIA Shaofeng. The quantitative analysis of factors influencing population distribution in China[J]. Geographical Research,1993,12(3):56-63.]

[5] 赵新正,李梦雪,李秋平,等. 中国副省级及以上特大城市人口空间分布与多中心性研究[J]. 干旱区地理,2017,40(2):415-423. [ZHAO Xinzheng,LI Mengxue,LI Qiuping,et al. Population distribution and polycentric spatial structure of Chinese megacities above the vice-provincial level[J]. Arid Land Geography,2017,40(2):415-423.]

[6] 王利,刘亚,曹文涛,等. 基于GWR模型的河北省县域人口分布的影响因素研究[J]. 辽宁师范大学学报(自然科学版),2016,39(2):262-267. [WANG Li,LIU Ya,CAO Wentao,et al. Research on influential factors of population distribution in Hebei Province based on GWR model[J]. Journal of Liaoning Normal University (Natural Science Edition),2016,39(2):262-267.]

[7] 童玉芬,马艳林. 城市人口空间分布格局影响因素研究——以北京为例[J]. 北京社会科学,2016,(1):89-97. [TONG Yufen,MA Yanlin. Influencing factors for the spatial distribution pattern of urban population:Take Beijing as an example[J]. Beijing Social Sciences,2016,(1):89-97.]

[8] 吕晨,樊杰,孙威. 基于ESDA的中国人口空间格局及影响因素研究[J]. 经济地理,2009,29(11):1797-1802. [LV Chen,FAN Jie,SUN Wei. Population distribution and the influencing fac-

ChinaXiv合作期刊

tors based on ESDA[J]. Economic Geography,2009,29(11):1797-1802.]

[9] 郭金铭,袁天凤,朱晓霞,等.川东北地区人口分布特征及影响因素分析[J].乐山师范学院学报,2013,28(8):100-103.[GOU Jinming,YUAN Tianfeng,ZHU Xiaoxia,et al. On the population distribution and influencing factors in northeast Sichuan[J]. Journal of Leshan Normal University,2013,28(8):100-103.]

[10] 沈思连,王春伟,汤静.基于GWR模型的河南省人口分布的影响因素研究[J].数学的实践与认识,2014,44(3):165-174.[SHEN Silian,WANG Chunwei,TANG Jing. Research on the influencing factors of population distribution in Henan Province based on GWR model[J]. Mathematics in Practice and Theory,2014,44(3):165-174.]

[11] 柏中强,王卷乐,杨雅萍,等.基于乡镇尺度的中国25省区人口分布特征及影响因素[J].地理学报,2015,70(8):1229-1242.[BAI Zhongqiang, WANG Juanle, YANG Yaping, et al. Characterizing spatial patterns of population distribution at township level across the 25 provinces in China[J]. Acta Geographic Sinica,2015,70(8):1229-1242.]

[12] 田盼盼,朱宇,林李月,等.省际与省内流动人口空间分布及其影响因素的差异——以福建省为例[J].人口学刊,2015,37(6):56-67.[TIAN Panpan,ZHU Yu,LIN Liyue,et al. Differences between the spatial distribution of floating population and its influencing factors between provinces and provinces:A case study of Fujian Province[J]. Population Journal,2015,37(6):56-67.]

[13] 汪思言,杨传国,庞华,等.珠江流域人口分布特征及其影响因素分析[J].中国人口.资源与环境,2014,24(5):447-450.[WANG Siyan,YANG Chuanguo,PANG Hua,et al. Characteristics of population distribution in the Pearl River Basin and the impact factors analysis[J]. China Population,Resources and Environment,2014,24(5):447-450.]

[14] 穆学英,刘凯,任建兰.新型城镇化背景下中国地级以上城市综合承载力空间格局研究[J].干旱区地理,2017,40(3):671-679.[MU Xueying,LIU Kai,REN Jianlan. Spatial pattern of the comprehensive carrying capacity of cities at prefectural level and above in China[J]. Arid Land Geography,2017,40(3):671-679.]

[15] 王桂新,潘泽瀚.我国流动人口的空间分布及其影响因素——基于第六次人口普查资料的分析[J].现代城市研究,2013,(3):4-11,32.[WANG Guixin,PAN Zehan. China's floating population spatial distribution and influencing factors:Evidence from year 2010 population census of China[J]. Modern Urban Research,2013,(3):4-11,32.]

[16] 张耀军,岑俏.中国人口空间流动格局与省际流动影响因素研究[J].人口研究,2014,38(5):54-71.[ZHANG Yaojun,CEN Qiao. Spatial pattern of population mobility and determinants of inter-provincial migration in China[J]. Population Research,2014,38(5):54-71.]

[17] 刘乃全,耿文才.上海市人口空间分布格局的演变及其影响因素分析——基于空间面板模型的实证研究[J].财经研究,2015,41(2):99-110.[LIU Naiquan,GENG Wencai. On evolution of the population spatial pattern in Shanghai and its influencing factors:Empirical research based on spatial panel model[J]. Journal of Finance and Economics,2015,41(2):99-110.]

[18] 王露,封志明,杨艳昭,等.2000—2010年中国不同地区人口密度变化及其影响因素[J].地理学报,2014,69(12):1790-1798.[WANG Lu, FENG Zhiming, YANG Yanzhao, et al. The change of population density and its influencing factors from 2000 to 2010 in China on county scale[J]. Acta Geographic Sinica,2014,69(12):1790-1798.]

[19] 张耀军,任正委.基于地理加权回归的山区人口分布影响因素实证研究——以贵州省毕节地区为例[J].人口研究,2012,36(4):53-63.[ZHANG Yaojun,REN Zhengwei. Factors affecting population distribution in mountainous areas: Geographically weighted regression using data from Bijie[J]. Population Research,2012,36(4):53-63.]

[20] 王周伟,崔百胜,张元庆.空间计量经济学现代模型与方法[M].北京:北京大学出版社,2017:1-31.[WANG Zhouwei, CUI Baisheng,ZHANG Yuanqing. Modern models and methods of spatial econometrics[M]. Beijing:Peking University Press,2017:1-31.]

[21] 张静,任志远.陕西省城市可持续发展系统协调性评价[J].地域研究与开发,2016,35(4):79-84.[ZHANG Jing, REN Zhiyuan. Coordination assessment of sustainable urban development system in Shaanxi Province[J]. Areal Research and Development,2016,35(4):79-84.]

[22] 石英,米瑞华.陕西省人口空间分异研究[J].干旱区地理,2015,38(2):368-376.[SHI Ying,MI Ruihua. Differentiation of population spatial distribution in Shaanxi Province[J]. Arid Land Geography,2015,38(2):368-376.]

[23] 陕西省地理国(省)情监测工作领导小组办公室.陕西地理省情白皮书(2012)[Z].[Office of the leading group for the monitoring of the situation of the geographical countries (provinces) in Shaanxi Province. Geography white paper of Shaanxi Province[Z].]

[24] 李仲生.人口经济学[M].2版.北京:清华大学出版社,2009:156-180.[LI Zhongsheng. Population economics[M]. 2nd ed. Beijing:Tsinghua University Press,2009:156-180.]

[25] 王少平,杨继生,欧阳志刚.计量经济学[M].北京:高等教育出版社,2011:61-113.[WANG Shaoping,YANG Jisheng,OUYANG Zhigang. Econometrics [M]. Beijing: Higher Education Press,2011:61-113.]

[26] BURRIDGE P. On the Cliff-Ord test for spatial correlation[J]. Journal of the Royal Statistical Society. Series B (Methodological),1980,42:107-108.

[27] ANSELIN L. Spatial econometrics:Methods and models[M]. Boston:Kluwer Academic Publishers,1988.

[28] BERA A K,MC ALEER M,PESARAN M H,et al. Joint tests of non-nested models and general error specifications[J]. Econometric Reviews. 1992,11:97-117.

[29] ANSELIN L. GIS research infrastructure for spatial analysis of real estate markets[J]. Journal of Housing Research,1998,9:113-133.

chinaXiv:202006.00008v1

ChinaXiv合作期刊

Factors influencing population distribution in Shaanxi Province using spatial econometric analysis

MI Rui-hua[1,2], GAO Xiang-dong[1]

(1 *School of Public Administration, Huadong Normal University, Shanghai* 260002, *China*;
2 *College of Economics and Management, Yan'an University, Yan'an* 716000, *Shaanxi, China*)

Abstract: When conducting empirical research on the factors influencing population distribution, it is helpful to understand population distribution and its evolutionary trends. This study considers all the 107 counties in Shaanxi Province, northwest China as research objects and attempts to use demographic, socioeconomic, and natural geographic statistical data for 2015 to fit an ordinary least squares (OLS) regression and spatial econometric model to analyze the factors influencing the Shaanxi Province's population distribution. The demographic and socioseconomic data were extracted from the 2016 Shaanxi Regional Statistical Yearbook, by the Statistics Bureau of Shaanxi Province in December 2016. The Shaanxi Province county-level administrative map was drawn using the National Earth System Science Data Sharing Infrastructure, National Science & Technology Infrastructure of China (http://www.geodata.cn). Natural geographical data were extracted, calculated, and analyzed from one kilometer (km) Resolution Digital Elevation Model Data Set of China from the Scientific Data Center of Cold and Arid Regions in China (http://westdc.westgis.ac.cn). The population spatial database was established using the ArcGIS 10.0 software program populated with the aforementioned data. The explanatory variable in our models is the Regional Proportion of Population (RPP), which is a proportion calculated by dividing the population of one county by the total population of the region. The RPP can effectively avoid the heteroscedasticity that may be caused by large differentials between the areas of the Shaanxi Province counties. The independent variables (economic and public service, population base, industrial structure, per capita Income, topography, and average elevation) were obtained through factor analysis to avoid overlooking variables and issues of multicollinearity. The OLS regression model demonstrates the fact that there is a significant relationship between RPP and the explanatory variables, as well as the dummy variables, defined in the model. The administration rank variable might play an important role in influencing the coefficient. However, the dependent variable and its residuals cannot satisfy the no spatial autocorrelation assumption, although they can satisfy the other Gauss-Markov assumptions. Therefore, we also attempt to fit the spatial lag model and the spatial error model (SEM). The SEM is the best model based on Lagrange multiplier (LM), Robust-LM, and Akaike information criterion tests. The SEM reveals that the RPP of a particular county in Shaanxi Province depends on not only the characteristics of observable variables but also other characteristics that may be unobservable or omitted from the model. Among the model's independent variables, the economic and public service factor exhibits the most significant positive explanatory power on the RPP. The population base factor, which represents basic agricultural conditions, and the population in the year 2000 demonstrate positive explanatory power. The industrial structure factor is negatively correlated with RPP, which indicates that the second industry has limited absorptive capacity in terms of increasing employment compared with the third industry. Average elevation has negative explanatory influence on RPP. The effects of per capita income and topography are insignificant, possibly because some counties with higher per capita incomes have economies based on natural resources or minerals and are often located in remote mountain areas. Nonetheless, topography exhibits a complex spatial coupling relationship with climate, precipitation, temperature, and humidity. The administration rank of a county influences population distribution significantly. Our main conclusion is that the key, controllable determinative factors for optimizing population distribution are socioeconomic factors, although the restrictive role of natural geographical factors should not be overlooked. By considering the spatial interactions between the explanatory variables and error terms, this study corrects the biased estimations of the OLS model and provides a scientific analysis of the factors influencing population distribution, which is of great reference value in terms of projecting as well as optimizing population distribution trends.

Key words: population distribution; influencing factors; regional proportion of population; spatial econometric model; Shaanxi Province

chinaXiv:202006.00008v1